United States
Department of
Agriculture

Forest Service

Intermountain
Research Station

Research Paper
INT-395

October 1988

Evaluating Statistical Techniques for Predicting and Interpreting FORPLAN Results

James F. C. Hyde III
J. Greg Jones
Mary L. Meacham

$$PNV = b_0 + b_1 X_1 + b_2 X_2 + \ldots + b_n X_n$$

THE AUTHORS

JAMES F. C. HYDE III is currently working as a consulting economist with R. W. Beck and Associates in Seattle. Prior to that he was an economist with the Economics Research Work Unit, Forestry Sciences Laboratory, Missoula, MT. He received his master's degree in economics from the University of Montana in 1985.

J. GREG JONES is a research forester with the Economics Research Work Unit, Forestry Sciences Laboratory, Missoula, MT. His research includes economic efficiency modeling for use in multiple-use management, measurement of multiple-use costs, and analysis of economic impact. He received his Ph.D. degree in forest economics from Iowa State University in 1976.

MARY L. MEACHAM is a statistician with the Economics Research Work Unit, Forestry Sciences Laboratory, Missoula, MT. Her research includes work in mathematical programming, regression analysis, and other statistical techniques. She is a graduate of the University of Montana.

ACKNOWLEDGMENTS

Data for this study were supplied by the Beaverhead, Gallatin, and Lolo National Forests. The planning staffs on these Forests were very helpful in retrieving and interpreting FORPLAN solution files. Several planning team members who played key roles deserve special thanks: Brad Burmark (Beaverhead NF), Michael Shaw (Gallatin NF), and Fred Stewart (Lolo NF). The authors also thank Michael Niccolucci, Intermountain Research Station, Missoula MT, for his thorough review and insightful suggestions, many of which have been incorporated into this paper.

Intermountain Research Station
324 25th Street
Ogden, UT 84401

RESEARCH SUMMARY

In the Forest Service's linear programming model, FORPLAN, many of the management objectives are expressed in the form of constraints that are difficult to evaluate in terms of their monetary effect on the objective function. Examples include forcing a certain amount of forage for wildlife, restrictions on sedimentation from roading activities, and the provision of recreational opportunities. These objectives all entail a cost in terms of output forgone as well as direct operating costs. Establishing the appropriate level of the constraint means comparing the benefits of including the constraint with these costs. The two approaches generally used to determine the costs are (1) the "shadow price" calculated in the linear programming process although this price is useful only for small incremental changes in the constraint and (2) solving the model twice. In the second approach the model is solved initially with the constraint set at the current level and then with it set to a substantially different level while all other things are held constant. The cost is the difference in present net value (PNV) between the two solutions. The disadvantages with this latter method are that it is both time consuming and very computationally intensive. Each constraint requires at least one solution, and because the constraints are generally not independent, but depend on the level and combination of other constraints imposed in the model, an even greater number of runs would be required to identify the costs of a large number of constraints across a variety of situations.

This paper explores another alternative, multiple regression analysis, for evaluating the cost of management constraints. Two general approaches were used to measure the effect of management constraints on PNV:

1. Ordinary least squares regression: The management constraints (formulated as independent variables) were regressed on PNV using a stepwise regression technique that tested all the variables at each step in the regression, bringing in those that are significant and rejecting those that become insignificant.
2. Principal component analysis combined with regression analysis: First, principal components comprised of management constraints were developed. Then, the principal components were regressed on PNV in the same stepwise manner as above.

These two approaches were tested on three National Forests in the Northern Region (Region 1). The data were taken from the FORPLAN solutions developed on the Beaverhead, Gallatin, and Lolo National Forests. A number of the original 33 solutions per Forest had to be dropped because they had different objective functions or used different cost or price data. The result was a data base for each Forest containing about 25 solutions. The constraints formed the independent variables. Some constraints were measured by continuous variables. Other constraints were measured by dummy variables, with 0 representing the absence of the constraint and 1 the presence of the constraint.

The highly interrelated nature of the constraints made it very difficult to develop regression models that included more than three or four constraints in the ordinary least squares approach. This impaired the predictive ability of the model because of the restricted number of variables in the equations. It was also not useful for interpretation because the constraints included in the equations served as proxies for a large number of interrelated variables. The coefficients, therefore, represented the amalgamation of those influences, not the single influence of the identified constraint. The second method overcame the first of these two obstacles by creating principal components that were composed of most of the management constraints and regressing them on the dependent variable, PNV. The advantage of principal components combined with regression over regression analysis lies in its ability to handle large data sets with many interrelated variables.

The principal component regression did include many more constraints of interest, and slightly enhanced the ability to predict what a given set of constraints would do to PNV. But it obscured even more the interpretability of the coefficients in terms of the effects individual management constraints have on PNV. The predictive ability is useful in those circumstances when the interrelationships inherent in the constraints being explored are the same as those interrelationships in the data used to identify the equations. Given the complexity of these relationships, it would be difficult to place much faith in the results of such a prediction.

The final problem, which neither approach was able to overcome, was the disparity between the number of constraints in each solution being very large, and the number of available runs in the data base used to develop each equation being very small. When there are a large number of variables relative to the number of variables to be identified, there is bound to be some difficulty in correctly evaluating the coefficients in the regressions.

As a result of these problems neither regression approach can be recommended as a satisfactory solution to the problem of identifying the cost of management constraints imposed in FORPLAN linear programming problems.

Evaluating Statistical Techniques for Predicting and Interpreting FORPLAN Results

James F. C. Hyde III
J. Greg Jones
Mary L. Meacham

INTRODUCTION

Many management objectives cannot be adequately reflected in terms of monetary value in the objective function in the linear programming formulations developed in FORPLAN. Examples include maintaining local economic stability and providing nonmarket benefits such as recreation or wildlife for which there are no suitable valuation methods. Objectives such as these are generally entered in the form of constraints. These management objectives, hereafter called constraints, can be expensive, both in terms of increased direct operating costs (Kemper and Davis 1976) and in outputs forgone (Bell and Randall 1982; Fight and others 1978). Identifying the appropriate constraints and levels thereof involves weighing the benefits of including the constraints against their costs. The approach recommended by the Secretary's Office, U.S. Department of Agriculture (MacCleery 1982) and suggested by numerous authors (for example, Fight and Randall 1980; Hughes 1965; Jones and others 1978) is to estimate the cost of a constraint in terms of the quantifiable outputs impacted, measured in terms of reduction in present net value or some analogous measure. A subjective judgment of the value of the constraint (because its value cannot be quantified) is then compared to the quantified cost to evaluate the constraint. Obviously, the constraint is desirable if its benefits are perceived to exceed its cost.

There are two common approaches for estimating the cost of constraints in linear programming models. One is the "shadow price" that is calculated for each constraint in the process of solving the linear program. The shadow price for a constraint measures the change in the objective function associated with changing that constraint by one unit, all else remaining constant. This cost, however, holds only over very small changes in the constraint. If the cost of a larger increment in a constraint is desired, the second approach for estimating cost is used. In this approach the model is solved twice, once with the constraint present at the current level and once with it set to a substantially different level, all else held constant. The cost of the constraint is the difference in the objective function values between the two solutions.

If the effects of a constraint on the objective function were independent of the other constraints imposed, the cost would only have to be calculated once for each constraint, since it would be constant over all situations. Experience has shown, however, that the cost of a constraint is seldom independent, but rather depends on the level and combination of other constraints imposed in the model. As a result, a large number of FORPLAN solutions would be required to identify the cost of a number of constraints over a variety of situations. FORPLAN solutions are expensive (often in the hundreds of dollars), yet the cost of constraints is needed to provide reasonable assurance that appropriate constraints are imposed at appropriate levels for each management alternative identified in forest planning. A less expensive method for estimating the cost of management constraints would be useful.

Multiple regression analysis could provide a third method for evaluating these costs. It involves applying multiple regression analysis to the results generated by FORPLAN. Equations could be developed that relate a dependent variable measuring present net value (or other objective functions), to independent variables measuring the extent to which constraints are imposed in FORPLAN solutions (acres of partial retention visual management, animal unit months of big game winter range, and so forth).

Such regression equations could serve two separate but related purposes. First, they could be used to predict values for the dependent variable given values of the independent constraint variables in the equation. Second, such regression equations could be used for explanatory purposes. In this use, the regression coefficients would measure the relationship exhibited between the dependent and independent variables.

If this application of regression proved successful, it would be useful in a number of ways during the latter stages of planning. For example, such equations could be used as an aid in "fine-tuning" management alternatives in forest planning. In some instances it may be desirable to relax certain constraints if such a change would save a large incremental cost. In other cases, it may be found that additional constraints could be imposed with little or no added cost. In addition, the procedure could be used to predict the cost of changes in alternatives proposed by the public in the review process of forest planning. This would aid Forest Service planners and decision makers in evaluating such suggestions and could provide some documentation underlying the ultimate decisions.

This paper reports the results of testing two approaches in which multiple regression techniques are used to fit present net value as a function of constraint variables. In one approach, called the ordinary least squares (OLS) approach, the constraint variables themselves are the independent variables in the regression model. In the second approach, principal component analysis with

regression (PCAR), principal components are computed from the original set of constraint variables. Then, multiple regression is used to fit present net value as a function of these principal components.

METHODS

The experiment was conducted using planning data from three National Forests in the Northern Region (Region 1): the Beaverhead, the Gallatin, and the Lolo. These Forests were selected because they were among those forests far enough along in the planning process to provide suitable data at the time the study began. Separate regression equations were developed for each Forest.

The FORPLAN solutions made in the forest planning process on these Forests served as the data base. Solutions were generated in two stages of planning. The first stage, called the analysis of the management situation, identified the range of outputs possible on the Forest, thereby defining the decision space for forest planning. Solutions generated in this stage vary substantially in terms of constraints imposed. In the second stage, management alternatives are developed. Generally, at least several solutions are made in the development of each management alternative, with a total of 20 to 40 solutions produced by this stage.

Our study began with approximately 33 solutions per Forest. It was necessary, however, to discard some of the FORPLAN solutions in order to maintain a consistent data base within each Forest. The solutions discarded include solutions with objective functions other than PNV and solutions that had different cost or price data, or a different base year for expressing constant dollars. The resulting data base consisted of 26 solutions for the Beaverhead National Forest, 25 solutions for the Gallatin National Forest, and 23 solutions for the Lolo National Forest.

The observed values for the dependent variable were the PNV objective function values (measured in thousands of dollars) obtained in the FORPLAN solutions. The independent variables measured the constraints imposed. In each Forest more than 100 such variables were developed. Some constraints were quantified as continuous variables. Examples include constraints placed on the volume of timber to be harvested in a particular decade, the acres of land to be allocated to specified uses such as wilderness and big game winter range, and acres to be managed with a high visual quality objective. For these, the right-hand-side value became the observed value. Other constraints, such as the nondeclining yield constraints, were quantified as dummy variables, where a value of 1 indicated the presence of the constraint at a given level and 0 the absence.

Many constraints measured as continuous variables were missing in at least several solutions. This posed no problem for lower limit constraints; for these variables a value of 0 was used when a constraint was not present (a lower limit of 0 is in fact equivalent to no constraint if the coefficients in the constraint are positive). A problem, however, arose in coding upper limit and equality constraints, which were quantified as continuous variables. Missing observations in equality and upper limit constraints could not be recorded as 0 when not imposed, since a right-hand-side value of 0 for an upper limit or equality constraint is extremely constraining. Treating these solutions as missing observations was not a viable option either, because these were valid data points even though some constraints were not imposed. Furthermore, treating them as missing data observations would reduce the number of usable observations to an unacceptably low level. This problem was handled as follows. Upper limit or equality constraints that were not present in all solutions were transformed to $1/C_j$, where C_j represents the right-hand-side value of the constraint in solution j. A value of 1 was used when an upper limit or equality constraint was not included in a solution. This transformation produced a variable whose values ranged from 0 to 1 (all values for C_j exceeded 1), thus avoiding the problem of a discontinuous variable or missing observations.

Early on it became obvious there was a high degree of collinearity among the constraints. Exact collinearity occurs if one variable is a linear combination of one or more of the other variables in the independent data set. As an example, two variables—X_1 and X_2—are exactly collinear if there exists a linear relationship, such as $X_2 = b_0 + b_1(X_1)$, which is true for all cases in the data (Weisberg 1980). In practice, such "exact collinearity" rarely occurs, and it is certainly not necessary for collinearity problems to exist (Belsley and others 1980). When two or more independent variables are collinear, the standard errors of the regression coefficients can become large. The estimates of the coefficients may be biased and the tests of significance become undependable due to the inflated standard errors.

A second but related problem is associated with measuring the effects of more than 100 independent variables with only 20 to 30 observations. This small number of observations limits the number of independent variables that can be identified in any one regression equation. Mathematically, the number of independent variables in any one regression cannot exceed the number of observations. But the number of independent variables required for statistical reliability is considerably less. Draper and Smith (1981) present the rule of thumb that there should be about 10 complete observations (degrees of freedom) for each variable to be estimated.

These problems made it necessary to reduce the number of independent variables in the data sets. The objective of the variable reduction process was to obtain a manageable list of constraint variables with correlations among themselves that were sufficiently low so as to avoid problems with collinearity and to try to reduce the disparity between the number of observations and the number of explanatory variables.

The first step was to eliminate the variables that were present in only several solutions as well as those variables that had the same value in each solution. It was apparent that these variables would not contribute to explaining variation in PNV.

In the next step, constraint variables were eliminated systematically so that no two variables in the remaining list of candidate variables had correlation coefficients of 0.9 or above. This was accomplished by listing each of the constraint variables in descending order according to their simple correlation with PNV. Each constraint variable

having a correlation of 0.9 or above with the first variable on the list was eliminated. Then, each constraint variable having a correlation of 0.9 or higher with the next (remaining) constraint variable on the list was exhausted. This process was continued until the entire list was processed. The result was a list of candidate constraint variables having the highest possible correlation with PNV and still having simple correlation coefficients among themselves less than 0.9.

The next step was to identify the extent of collinearity among the remaining independent variables. This was accomplished by calculating a variance inflation factor for each variable. The variance inflation factors are calculated as the diagonal elements of the inverse of the correlation matrix. A high variance inflation factor (say a value greater than 10) indicates that a high degree of collinearity exists among the independent variables. It, however, does not indicate which variables are the cause of this collinearity. (Readers are referred to Belsley and others 1980 for a complete discussion of variance inflation factors as well as other techniques for handling collinearity.)

If the presence of high collinearity was indicated by the variance inflation factors, the list of candidate constraint variables was pared down further. This second stage of variable reduction used a series of diagnostic regressions. In this process, each constraint variable was regressed against the set of remaining independent variables using stepwise regression procedures. Each of the diagnostic equations having adjusted R-squares of 0.9 and above was examined, and the dependent variable in that equation was deleted from the list of candidate variables. Eliminating this variable eliminates the source of collinearity between it and the significant independent variables in the equation (they are highly related as indicated by the high R-square), and yet most of the variation in the eliminated variable can still be represented by the remaining independent variables, which are not highly correlated among themselves. New variance inflation factors were calculated for the resulting list of candidate constraint variables. If collinearity was still present, this second stage of variable reduction was repeated until the collinearity was reduced to an acceptable level. The result of this process was a set of candidate variables that were used in an ordinary least squares regression modeling procedure.

The regression equations were developed using a stepwise regression approach in which variables enter the model one at a time in the order of their explanatory strength. At each step in the procedure, the variables currently in the model are tested for significance, and insignificant variables are removed.

OLS RESULTS AND DISCUSSION

The regression equations developed by fitting PNV as a function of constraint variables are presented in tables 1, 2, and 3. Both unstandardized regression coefficients (measured in the original units) and standardized regression coefficients (transformed to 0 mean and standard deviation of 1) are presented. Two of the three models represent reasonably good fits by standard statistical measures. These are the Beaverhead model (table 1), which contained

Table 1—Equations developed for the Beaverhead National Forest by fitting PNV as a function of constraints

Variables in equation[1]	Unstandardized coefficients	Standardized coefficients
LIMA1	41,800	0.375
P41S3	36,200	.653
P16S2	421	.299
P48S4	38,200	.828
(Constant)	–74,300	
R-square = 0.77	Adjusted R-square = 0.72	F = 15.4

[1]LIMA1 = upper-limit constraint placed on volume of timber that can be harvested in the Lima Mountain Range in decade 1, measured in millions of cubic feet.

P41S3 = the sum of the equality constraints that force water rehabilitation prescriptions to be allocated on a given number of acres across the Forest (measured in thousands of acres).

P16S2 = the sum of the equality constraints that force timber management prescriptions of medium intensity, visual quality objectives of modification or maximum modification, to be allocated on a given number of acres across the Forest (measured in thousands of acres).

P48S4 = the sum of the equality constraints that force nontimber summer range prescriptions for big game to be allocated on a given number of acres across the Forest (measured in thousands of acres).

Table 2—Equations developed for the Gallatin National Forest by fitting PNV as a function of constraints

Variables in equation	Unstandardized coefficients	Standardized coefficients
V16A04[1]	365	–0.470
(Constant)	297,000	
R-square = 0.22	Adjusted R-square = 0.19	F = 6.5

[1]V16A04 = the sum over mountain ranges of the upper-limit constraints placed on the volume of timber that can be harvested in decade 1, measured in millions of cubic feet.

Table 3—Equations developed for the Lolo National Forest by fitting PNV as a function of constraints

Variables in equation[1]	Unstandardized coefficients	Standardized coefficients
MATOLDLL	–0.792	–0.355
TM812	–29,200	–.366
WFOR1LL	–120	–.348
(Constant)	413,000	
R-square = 0.82	Adjusted R-square = 0.80	F = 29.9

[1]MATOLDLL = a lower-limit constraint placed on the mature old growth provided for wildlife species that are mature old-growth users, measured in thousands of acre equivalents.

TM812 = a dummy variable identifying when equality constraints were applied to force the allocation of specified amounts of grizzly bear habitat management, coded as 1 when these constraints were present and 0 when they were absent.

WFOR1LL = lower-limit constraint placed on the amount of winter forage for big game that is to be produced for decade 1, measured in thousands of animal unit months (AUM's) per decade.

four constraint variables explaining 77 percent of the variation in PNV, and the Lolo model (table 2), which included three constraint variables explaining 82 percent of the variation in PNV.

In the case of the Gallatin model (table 3), only one constraint variable was significant at the 10 percent level, and the regression equation explained 22 percent of the total variation in PNV. A major reason for such a poor fit was that few of the constraints were present in even half of the solutions, and thus the degree of variation in the explanatory variables was greatly reduced.

Let us now consider these models in view of one of the previously stated uses, that of interpretation. In this use, the regression coefficients are used as a measure of the per-unit effect the associated independent variable (constraint) has on the dependent variable (PNV). For example, in the Beaverhead equation, the unstandardized coefficient for LIMA1 (table 1) indicates that PNV would increase $41.8 million for a one-unit increase in LIMA1. The standardized coefficient can be used to examine the strength of the relationship. In this case the coefficient for LIMA1 indicates that it is the third most important variable in the equation.

There are, however, two major problems with using the regression coefficients in any of these models for interpretation purposes. First, recall that many of the original constraint variables were dropped because they were highly correlated with one or, in most cases, many of the other independent variables. (The problem of collinearity discussed earlier.) As a result, it is likely that the variables included in the above model are serving as proxies for the highly collinear variables that were dropped. Therefore, the regression coefficient of a particular variable measures not only the effects of that variable, but the joint effect of the set of correlated variables which it represents.

The second difficulty with using the regression coefficients for interpretation is that unless the independent variables included in a model are strictly independent (are not correlated with each other), the regression coefficient for any given variable may be influenced by which additional variables are included in the regression model. Therefore, to use the regression coefficients for interpretation purposes, one must assume that the model is correctly specified and make a judgment about the extent to which collinearity may influence the estimation of the coefficients. Given the fact that there are a large number of interrelated management constraints used in FORPLAN, and that these regression models include at most four variables, the probability that these regression models are correctly specified seems quite small. For this reason, plus the fact that the variables in the models are acting as proxy variables, these models appear to have little value for measuring the effect of specific constraints on PNV.

Prediction was the other stated use for these regression equations. Prediction has a distinctly different focus than does interpretation. Although interpretation uses the regression coefficients to quantify the relationship between a dependent and an independent variable, prediction is concerned with using the independent variables to predict values for the dependent variable. Collinearity and model misspecification are much less of a concern in prediction, because good predictions can still be made even when these conditions are present (Koutsoyiannis 1979). A major concern in prediction is that the sets of values for the independent variables from which PNV is predicted exhibit the same relationships that were present among these variables in the data from which the equations were developed. In other words, these equations should only be used to predict PNV for the types of management alternatives included in the original data. Given the complexity of each management alternative (hundreds of constraints), it is unlikely that the user would be able to ascertain whether a particular management alternative was similar to those used to develop the equations. Large errors are possible if PNV is predicted for a vastly different type of management alternative.

In most prediction applications, it is desirable to develop regression equations that reliably predict the dependent variable with as few independent variables as possible. This minimizes the amount of data that must be collected both for fitting the equation and applying the equation to predict the dependent variable. The situation here, as it turns out, is somewhat different. We would prefer more rather than fewer independent variables in a prediction equation, because this would allow us to predict PNV for more precisely defined management alternatives. As it is, PNV can be predicted only from four constraints in the Beaverhead model, one in the Gallatin model, and three in the Lolo model. It is not possible to predict PNV for a management alternative that involves specifying one or more constraints that are not present in these models, and this restricts the usefulness of these models for predictive purposes.

These regression models are constructed using the criterion that an independent variable had to be statistically significant at the 10 percent confidence level to be included in the regression equations. There is a school of thought that these criteria should be relaxed to much lower levels if a regression model is to be used solely for prediction purposes. It is argued that this would increase the number of independent variables included in the model and thus the accuracy of the fit. It is apparent, however, that the increases in the number of variables that are possible would only marginally improve the previously discussed problem. The number of variables permitted in a regression model is ultimately limited by the number of observations available in the original data. The number of observations (FORPLAN solutions) from which these equations were developed would have to be increased substantially before much improvement in this problem could be realized. To illustrate, say that there are 50 constraints of management importance (probably somewhat conservative based on the FORPLAN models we have seen). If the same ratio of number of independent variables to number of observations is maintained at about 10 observations per independent variable, 500 FORPLAN solutions would be required for a regression model to contain all 50 variables. This would be difficult, given the cost of producing those solutions.

THE PRINCIPAL COMPONENTS/ REGRESSION APPROACH

The problem, mentioned above, of being able to predict PNV based on only a small number of constraint variables may be solved to some extent through the use of principal components. In this approach, a principal component analysis would be performed on the complete set of constraint variables of interest. Then, a regression analysis would be done using the principal components as independent variables.

The objective of principal components analysis is to transform a set of variables into a new set of variables that are not correlated and capture as much of the variance of the original variables as possible. The first principal component is the linear combination of all of the variables which captures the greatest amount of variation in the data set. The second component is the linear combination of all of the variables, which has as much of the remaining variance as possible subject to being orthogonal (linearly unrelated) to the first and so on for the rest of the principal components.

Figure 1 illustrates the concept of principal components. X_1 and X_2 are two hypothetical variables. When plotted, the distribution of the observed values takes the form of an ellipse (fig. 1a). If a principal component analysis were to be performed, the first principal would be along the new axis labeled Y_1 (fig. 1b). Y_1 could be thought of as a variable comprised of both X_1 and X_2 (it could be represented mathematically as a function of X_1 and X_2). This dimension explains (or measures) the greatest amount of variation possible for a single dimension (the range of the distribution is greater along this new axis Y_1 than any other axis that could be drawn).

The second principal component in this example would be the new axis Y_2, which is orthogonal to Y_1. The distribution and dimensions Y_1 and Y_2 can be redrawn as illustrated in figure 1c, because Y_1 and Y_2 represent linear transformation of the original variables X_1 and X_2. Together, Y_1 and Y_2 account for 100 percent of the variation in the original variables, X_1 and X_2.

The advantage of the principal components combined with regression over classical regression analysis is its greater facility for handling large, highly interrelated sets of data. Because the principal components are orthogonal, the information provided by each is unique. Thus, the components are not linearly related, so the problem of collinearity in the regression portion of the analysis is eliminated. When the matrix of independent variables contains some variables that are perfect linear combinations of others (collinear), the regression problem cannot be solved, but by using the principal components regression technique, it is possible to estimate the regression parameters (Massy 1965). The principal components provide the same total information as the original variables, and any analysis between the variables which compose the principal components and any other variables can use the principal components with no loss of information.

Just as the principal components are defined as linear combinations of the original variables, the original variables can also be defined as linear combinations of the principal components, and there is a unique relationship

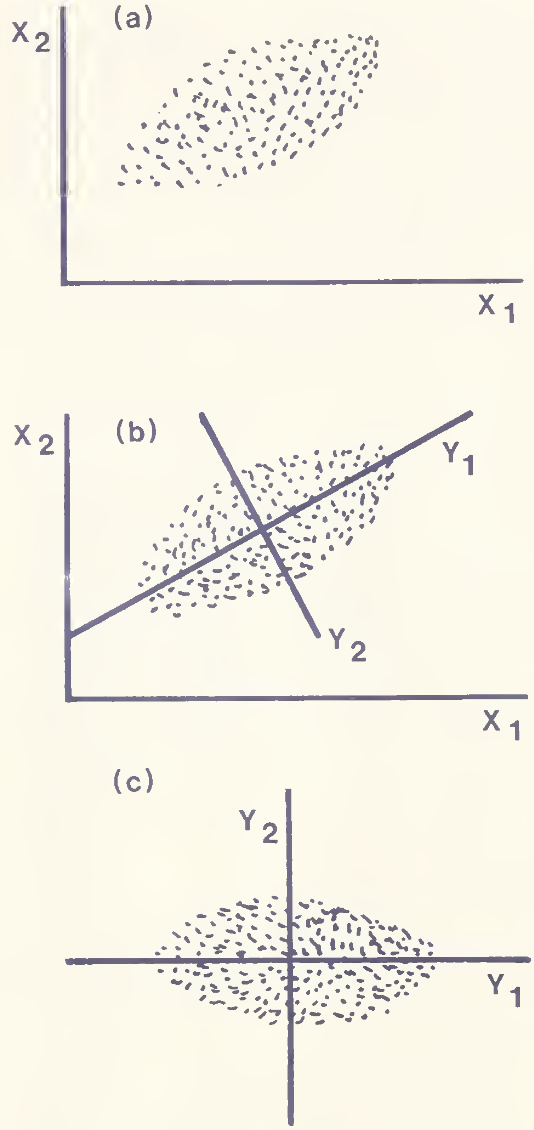

Figure 1—Example illustrating the basic concept underlying principal components: (a) is a plot of two hypothetical variables, X_1 and X_2; (b) Y_1 and Y_2 are the two principal components; and (c) shows rotating the axis so that Y_1 and Y_2 become, respectively, the ordinate and the abscissa.

between the coefficients on the principal components in the regression and the variables within the principal components. Thus the transformation of the regression coefficients back to the original variables yields the usual least squares solution for the beta coefficients of regression (Massy 1965). This entire process, however, has done nothing to reduce the inherent collinearity of the underlying data, and the resulting coefficients are still dependent on the variables included in the principal component analysis. What has improved is the accuracy with which the coefficients can be estimated by the regression procedure.

The first step in applying PCAR was to decide which constraint variables should be included in the principal

components. Recalling that there were about 100 constraint variables for each Forest, we excluded from the PCAR analysis those variables whose values had little to no variation across most FORPLAN solutions and variables which were not present in most solutions. The remaining variables (48 on the Beaverhead, 49 on the Lolo, and 15 on the Gallatin) were the variables potentially useful in explaining variation in PNV.

There are two bases for generating principal components, using the variance-covariance matrix and using the correlation matrix. We used the correlation matrix, which is more stable and should be used when the variables are measured in a variety of units such as occurred in this study. The other approach is sensitive to both units as well as scale and should be applied only when all the variables are measured the same.

Because initially there are as many principal components as original variables, not all of the principal components could be present simultaneously in a regression equation. Removing nonzero principal components reduces the amount of information to the regression. There are two criteria that can be used to determine which components to delete from the regression analysis (Massy 1965): (1) those principal components that have a low correlation with the dependent variable can be dropped, or (2) those principal components with the smallest eigenvalues (lowest variation) can be dropped. We used the latter criteria, dropping only those components with extremely small eigenvalues, so that as much of the variation in the original independent variables as possible was preserved in the principal components used.

One of the problems with principal components analysis is that there is no way of investigating whether a particular variable included in the PCAR is related to the dependent variable. And because using different numbers of variables or principal components gives substantially different results, it appears that including only relevant variables in the PCAR is very important. We excluded only variables that created problems in the data set, such as those that were constant, missing in most cases, or had very little variation across the data.

PCAR RESULTS AND DISCUSSION

The first step in the PCAR approach is to compute the principal components. Table 6 presents the eigenvectors for the 10 principal components computed for the Lolo (the constraint variables are defined in the appendix). These eigenvectors present each principal component as a function of the original constraint variables. The 10 principal components captured essentially all of the variation in the set of 49 constraint variables. In other words, these 10 principal components contain essentially the same information as do the original constraint variables.

Once the principal components were computed for each data point, the stepwise regression procedure used previously was applied to fit PNV as a function of the principal components. The model developed for the Lolo is presented in table 4. Three principal components significant at the 10 percent level are included in this model. The adjusted R-square for this model is 0.78.

A summary of the results from the PCAR approach is presented for all three Forests in table 5. As was the case in the OLS approach, a relatively good fit was obtained for the Lolo and Beaverhead cases (adjusted R-squares were 0.78 and 0.80, respectively), but a relatively poor fit was obtained for the Gallatin case (adjusted R-square of 0.20). A contributing factor on the Gallatin was that the two principal components in that model explained only 24 percent of the variation in the original constraint variables. For the Lolo and Beaverhead, the principal components included in the models captured much more variation in the original constraint variables, 82 and 83 percent, respectively.

In terms of the variation in PNV, the models containing the principal components as independent variables generally provided a modest improvement over the approach in which PNV was fit against the constraint variables themselves. The adjusted R-squares from both approaches are:

	OLS approach	PCAR approach
Beaverhead	0.72	0.80
Lolo	.80	.78
Gallatin	.19	.20

Table 4—Equations developed for the Lolo National Forest by fitting PNV as a function of principal components

Variables in equation[1]	Unstandardized coefficients	Standardized coefficients
PC1	−75.6	−0.643
PC3	−22.6	−.566
PC4	−10.8	−.270
R-square = 0.80	Adjusted R-square = 0.78	F = 27.7

[1]PC1, PC3, and PC4 = respectively, principal components 1, 2, and 4 that are calculated from the constraint variables as shown in table 6.

Table 5—Summary of results from fitting PNV as a function of principal components

Item	National Forest		
	Beaverhead	Lolo	Gallatin
R-square	0.92	0.81	0.26
Adjusted R-square	0.80	0.78	0.20
Number of constraint variables comprising the principal components	48	49	15
Number of principal components in the regression equations[1]	7	3	2
Percentage of the total variance in the original constraint variables accounted for by the principal components in the regression equations	82	83	24

[1]Principal components included in the regression equations are significant at the 10 percent confidence level.

The greatest improvement was observed for the Beaverhead, an increase of 8 percent; a 1 percent increase was observed for the Gallatin, and a slight drop occurred for the Lolo.

Interpreting the regression equation developed via the PCAR approach depends on one's ability to relate each principal component back to the original constraint variables. The key here is the coefficients in the eigenvectors that relate the constraint variables to the principal components (table 6).

If the coefficient of a constraint variable in the eigenvector is large, that variable strongly influences the principal component and is said to have a "high loading." When a principal component is composed mostly of variables of the same type of constraint, and the loadings of the others approach zero, then it is possible to interpret

Table 6—Conversion factors (called eigenvector elements) for computing the principal components for the original constraint variables for the Lolo National Forest

Constraint variables[1]	PC1	PC2	PC3	PC4	PC5
CV2	0.1627	0.0081	0.0020	–0.0880	0.1623
CV3	.1570	.0688	–.1184	.1059	–.0814
CV4	.1493	.1045	.1170	–.0481	–.2972
CV5	.0994	–.2971	–.0238	.0504	–.0720
CV6	.1574	.0694	–.1351	.0921	.0068
CV7	.1652	.0011	–.0348	.0028	.0099
CV8	.1590	–.0140	–.1361	.1184	–.0291
INVCV1	.1442	–.1795	–.0796	.0959	–.0418
DWHGEQ	.1540	.0884	.1054	–.0551	–.2412
INVAV4	.0359	.1219	.4021	–.4706	.1384
DWVAEQ	.1638	.0108	–.0770	.0236	.0627
INVKV16	.1589	–.0034	–.1388	.1191	–.0281
MATOLDLL	.0997	.0415	.3508	.3396	–.0717
INVCT21	.1132	.0213	.2835	.3285	–.0102
INVLN19	.1611	.0696	.0576	–.0662	–.0756
WFOR1LL	.0580	.0295	.3409	.4306	.1507
HCLL	.1461	.0995	.1369	–.0469	–.3509
INVNC11	.1088	–.1021	.3571	–.2514	–.2734
HGLL	.1536	.0872	.1094	–.0558	–.2495
RALL	.1640	–.0289	.0430	–.0682	.0107
NRARB	.1640	–.0291	.0430	–.0682	.0106
LINEFV18	.0885	–.3327	.0096	.0435	–.0463
RBLL	.0885	–.3327	.0096	.0435	–.0463
TBLL	.1196	.2509	–.1517	.0794	.0274
INVNHDL	.0885	–.3327	.0096	.0435	–.0463
INVHCL	.1442	–.1795	–.0796	.0959	–.0418
INVNH7	.1590	–.0143	–.1360	.1184	–.0291
INVIV19	–.0478	–.3469	–.1336	–.2154	.0107
INVEV19	.0208	–.3863	.0799	–.0416	–.0605
LINECV8	.1610	.0181	–.1045	.0966	–.0685
LINEEV20	.1231	.2471	–.1192	.0775	–.0555
LINEFV8	.1624	.0402	.0692	–.0730	–.0760
LINEAV8	.1635	.0263	–.0202	–.0589	.1319
DWRAEQ	.1615	.0424	–.1046	.0571	.0312
NHGMM	.1648	–.0046	.0156	–.0534	.0174
NHAVD	.1648	.0125	–.0028	–.0475	.0592
NHDUL	–.1551	.1256	.0106	–.0164	.1475
MNLL	.1632	–.0356	–.0168	–.0509	.1074
VCLL	.1637	–.0021	–.0134	–.0634	.1390
VDLL	.1630	.0035	–.0022	–.0807	.1552
TRELL	.1628	.0041	.0022	–.0871	.1609
VBLL	.1622	.0055	.0103	–.0988	.1713
NVAVD	.1623	.0060	.0079	–.0957	.1687
HBLL	.1633	.0110	–.0091	–.0734	.1498
VALL	.1613	.0130	.0181	–.1119	.1838
NHAHB	.1644	.0004	–.0474	–.0158	.0969
LEGHTV4	.1626	.0259	–.0069	–.0807	.1584
NUTBLL	.1625	.0464	–.0843	.0239	.0673
NDV2	–.0427	–.1248	.3465	.1828	.4357

(con.)

Table 6 (Con.)

Constraint variables[1]	PC6	PC7	PC8	PC9	PC10
CV2	0.0615	0.0093	–0.0379	0.0873	–0.0323
CV3	–.0951	–.0085	.1126	–.2259	.0533
CV4	–.1558	–.0656	–.0131	.0413	.0077
CV5	.0627	–.0596	–.5345	–.2419	.6687
CV6	–.0564	–.0014	.0608	–.0902	.0039
CV7	–.0202	–.0031	.0209	–.0065	–.0254
CV8	–.0580	.0052	.0894	–.1216	–.0189
INVCV1	–.0177	.0196	.0992	–.1086	–.0683
DWHGEQ	–.1282	–.0563	–.0162	.0486	.0008
INVAV4	.1139	.0338	.3551	–.6359	.1535
DWVAEQ	–.0054	.0060	.0291	–.0248	–.0212
INVKV16	–.0602	.0043	.0882	–.1217	–.0157
MATOLDLL	.2862	.3268	–.3470	–.2500	–.3758
INVCT21	.1316	.4857	.4340	.3525	.4821
INVLN19	–.0546	–.0308	–.0254	.0628	–.0087
WFOR1LL	.2332	–.7687	.1514	.0346	.0435
HCLL	–.1764	–.0728	–.0100	.0383	.0075
INVNC11	–.0463	–.0597	–.1044	.2310	–.0653
HGLL	–.1311	–.0574	–.0162	.0490	.0006
RALL	.0034	–.0073	–.0153	.0616	–.0402
NRARB	.0035	–.0073	–.0152	.0616	–.0402
LINEFV18	.0357	.0318	.0852	–.0644	–.1102
RBLL	.0357	.0318	.0852	–.0644	–.1102
TBLL	–.0818	–.0185	.0221	–.0662	.0633
INVNHDL	.0357	.0318	.0852	–.0644	–.1102
INVHCL	–.0177	.0196	.0992	–.1086	–.0683
INVNH7	–.0579	.0052	.0894	–.1216	–.0190
INVIV19	–.1069	–.1651	.2474	.1430	.1232
INVEV19	.0472	.0103	.0766	.0139	–.1287
LINECV8	–.0769	–.0066	.0741	–.0995	–.0109
LINEEV20	–.1148	–.0309	.0251	–.0669	.0625
LINEFV8	–.0467	–.0284	–.0248	.0678	–.0184
LINEAV8	.0382	.0068	–.0140	.0329	–.0138
DWRAEQ	–.0328	.0019	.0484	–.0685	–.0036
NHGMM	.0070	–.0244	–.1363	.0129	.1247
NHAVD	.0105	–.0084	–.0600	.1162	–.0294
NHDUL	.0815	.0293	–.0843	.0359	.0354
MNLL	.0537	–.0124	–.1726	.0002	.1627
VCLL	.0486	.0095	–.0214	.0619	–.0332
VDLL	.0580	.0095	–.0328	.0797	–.0330
TRELL	.0617	.0096	–.0368	.0862	–.0334
VBLL	.0684	.0098	–.0441	.0980	–.0340
NVAVD	.0664	.0097	–.0422	.0949	–.0335
HBLL	.0524	.0086	–.0295	.0727	–.0300
VALL	.0748	.0095	–.0531	.1116	–.0327
NHAHB	.0199	.0080	.0069	.0142	–.0281
LEGHTV4	.0542	.0075	–.0362	.0810	–.0262
NUTBLL	–.0116	.0028	.0237	–.0229	–.0103
NDV2	–.7873	.0700	–.1061	–.0507	–.0267

[1]The constraint variables are defined in the appendix.

the meaning of that principal component. For example, a principal component comprised of a high loading for a constraint variable measuring lower limits placed on timber harvesting and a high loading for a constraint that represents lower limits placed on forage production (with all other variables having low loadings) could be thought of as representing intensive forest management. The regression coefficient associated with that principal component would measure the relative effect of intensive management on PNV.

But high loadings on discernible groups of constraint variables with low loadings on others did not occur with any frequency on the three Forests tested in this study. The eigenvectors in table 6 illustrate quite well the patterns observed in this study—a range of loadings, some large, some medium, and some small, with no easily

identifiable pattern. As a result it is difficult to attach any interpretive meaning to the principal component obtained.

To predict PNV for a management alternative, one would first define the management alternative in terms of all the constraints to be imposed. These values would then be substituted into the eigenvector equations (shown for the Lolo in table 6) to compute the values for the principal components. The values for the principal components are then substituted into the regression equations to compute the predicted PNV.

The principal component approach provides the ability to predict PNV for much more precisely defined management alternatives (in terms of the constraints included) than does the OLS approach using the original variables. For example, on the Lolo, 49 constraint variables can be used to define a management alternative with the PCAR approach, while using the OLS approach a management alternative could only be defined using three constraint variables. One of the major restrictions that was discussed previously regarding the use of the OLS equations for predicting PNV also applies to the principal components equations. These equations should be used only to predict PNV for management alternatives that are similar to the alternatives (FORPLAN solutions) that comprised the data from which the equations were developed. In other words, the relationship between the constraints in a trial management alternative must be similar to those in the data used to develop the equations. A problem with this restriction is that there are so many dimensions to a typical planning problem that it is difficult to know when one is attempting to predict PNV for a management problem that is out of range for a prediction equation.

It seems that the interrelated nature of constraints that is present in linear programming models such as FORPLAN is a major problem underlying the general approach tested in this study. The effect of any one constraint on PNV depends on the level at which other constraints are imposed. Perhaps this point is best explained via a simple example. Figure 2 represents a simple linear programming problem having two outputs, X_1 and X_2. Two upper-limit constraints are imposed, C_1 and C_2, resulting in a feasible region OCBA. Assume now that an additional upper limit constraint, C_4, is imposed (fig. 3) and that the objective function has a slope that is less than the slope of C_4 but greater than the slope of C_1 (line BA in fig. 3). Then, the largest feasible value for the objective function would occur at the corner point Y_1. If constraint C_4 is relaxed (increased) to C_4', the optimal solution would instead occur at Y_2, and if moved to C_4'', would occur at A. But relaxing C_4 further so that it intersects the horizontal axis to the right of point A (as illustrated by C_4''' and C_4'''') would have no additional effect on the optimal solution. It would continue to stay at point A, where constraint C_1 is binding.

The optimal solution points identified in figure 3 are plotted against PNV in figure 4. The regression line that would result from regressing PNV against C_4 is included. Note that the relationship between C_4 and PNV is not linear. The rate of change in PNV associated with a unit change in C_4 depends on what other constraints are imposed. As a result, the regression line underestimates PNV in the middle of the distribution and overestimates

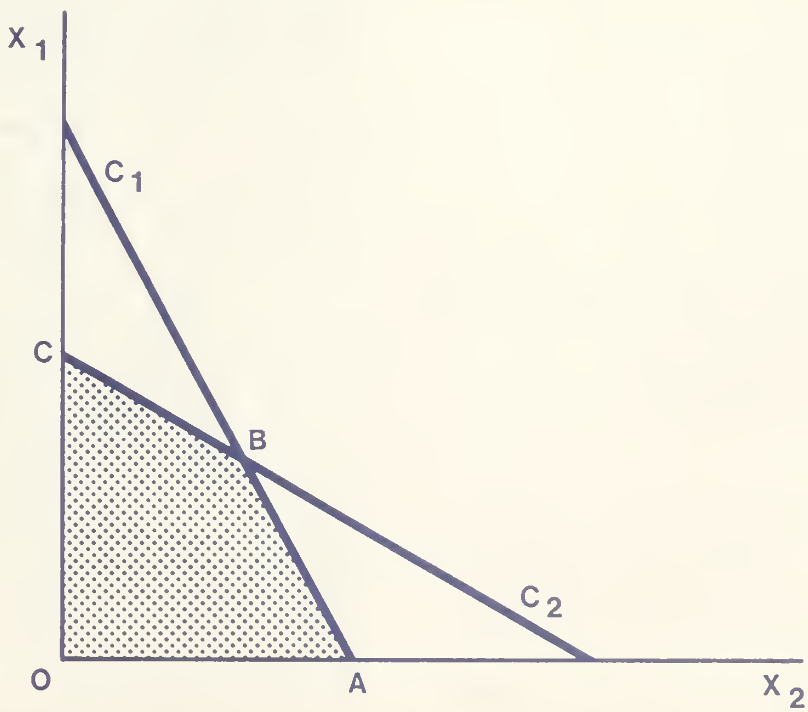

Figure 2—A hypothetical linear programming problem containing two output variables, X_1 and X_2, two upper-limit constraints, C_1 and C_2, and feasible region OCBA.

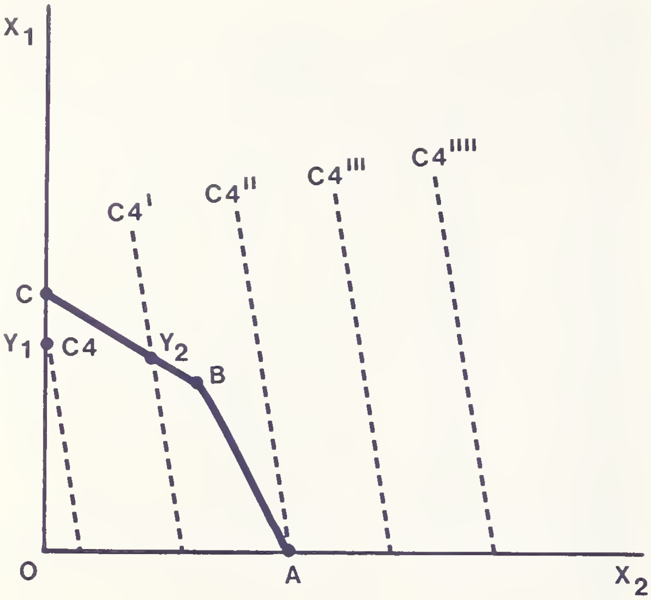

Figure 3—A series of hypothetical linear programming problems in which upper-limit constraint C_4 is systematically relaxed from C_4 to C_4''''.

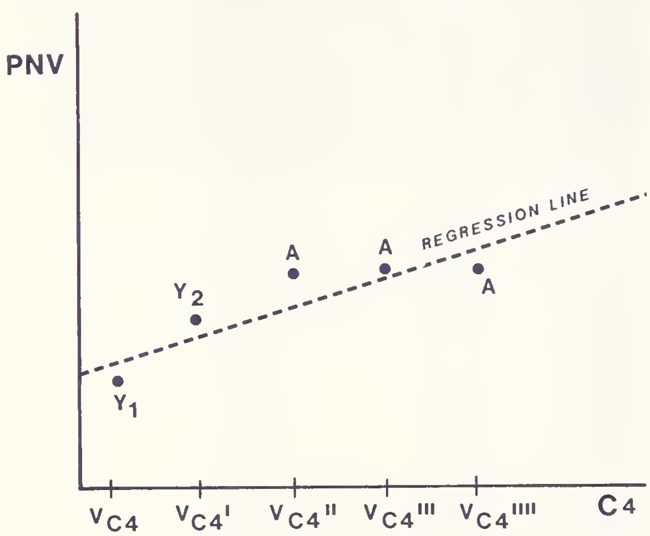

Figure 4—The right-hand-side values for the C_4 series of constraints (fig. 3) plotted against the maximum values for PNV.

on the ends. But there are worse scenarios. If C_1 and C_2 (fig. 2) were increased or decreased, or additional constraints were imposed, the relationship between PNV and C_4 in figure 4 could take on an entirely different shape. Again, the effect C_4 has on PNV depends on what other constraints are imposed.

It is believed that relationships such as these contribute to the large amount of collinearity observed in the test data sets. The result is difficulty in isolating the effect of any one constraint on PNV because that effect is not independent of a large number of other circumstances. These relationships also can cause problems in predicting PNV. Consider again the example presented in figures 2 through 4. Because the relationship between PNV and C_4 depends on what other constraints are imposed, and at what level they are imposed, any number of quite different regression lines could be obtained for PNV and C_4. In other words, the results could vary substantially, depending on what solutions comprise the data base. Furthermore, this problem is exacerbated by the fact that only a relatively small number of solutions would likely serve as the data base for the analysis. This emphasizes the importance of the statement made earlier that PNV should be predicted only for the same types of management alternatives present in the data used to develop a prediction equation.

CONCLUSIONS

The results indicate that the two approaches tested in this study do not reliably estimate the effect individual constraints have on PNV—the interpretative use of the regression coefficients. With the OLS approach, inherent collinearity among the constraints makes it difficult to correctly isolate the effect of any one constraint. Using principal component analysis with regression, relationships between the principal components and PNV can be accurately determined, but it is difficult to relate this information back to the original constraints.

For predictive purposes, these approaches fared only slightly better. With the OLS approach it was possible to have only a small fraction of the constraints in a predictive equation simultaneously. This severely limits their usefulness because they cannot be used directly to predict PNV for management alternatives that involve specifying constraints not present in the equations. The principal component approach improved this situation somewhat. Two of the three test models contain principal components comprised of about 50 constraints. This allows a much more precisely defined management alternative to be evaluated.

An important restriction, however, is present when using either OLS or principal component equations for prediction. The effect of any one constraint on PNV is generally determined by what other constraints are imposed. Thus, the equations should only be applied to predicting PNV for the same "types" of management alternatives that were present in the data used to develop those equations. Substantial errors could occur if PNV were to be predicted for a management alternative that involved a pattern of constraints that was not present in the original data. Given the many dimensions of a typical planning problem, this is a serious limitation. It is likely that it would be extremely difficult to identify when one is attempting to predict PNV for a management alternative (set of constraints) that contains substantially different constraint interrelations than the data used in developing the prediction equation.

A further problem affecting both the OLS and principal component approaches was the limited number of observations in the data. When there is a large number of variables relative to the number of observations, some of the variables are bound to be collinear just by chance.

Thus, collinearity problems can be expected to occur in sample data even when collinearity is not present in the population being sampled. Second, because the number of data points is small, only a small portion of the decision space is represented in the data. Moreover, there is reason to believe that the data points sampled are not well distributed across the decision space. Sets of very similar solutions were typically made in developing a forest plan alternative, giving rise to clusters of data points. This restricts the range of management alternatives for which the equations can reliably predict PNV.

A further consideration is that the application of this approach requires a substantial amount of resources. A large quantity of information must be extracted from each FORPLAN solution and coded into workable constraint variables. In addition, a good knowledge of statistical model building techniques (particularly techniques for dealing with collinearity) is needed.

It is clear that a number of problems exist in the approaches tested in this study. As a result, we believe the methods have very limited applicability. In view of this, and in view of the resources and expertise required, the approaches of fitting PNV as a function of constraints either directly with OLS or indirectly with principal components, cannot be recommended for either prediction or interpretation purposes.

REFERENCES

Bell, Enoch F.; Randall, Robert M. 1982. Opportunities to increase national forest softwood harvest. Journal of Forestry. 80: 152-156, 160.

Belsley, David A.; Kuh, Edwin; Welsch, Roy E. 1980. Regression diagnostics: identifying influential data and sources of collinearity. New York: John Wiley and Sons. 292 p.

Draper, N. R.; Smith, H. 1981. Applied regression analysis. New York: John Wiley and Sons. 709 p.

Fight, Roger D.; Randall, Robert M. 1980. Visual quality and the cost of growing timber. Journal of Forestry. 78: 546-548.

Fight, Roger D.; Johnson, K. Norman; Connaughton, Kent P.; Sassaman, Robert W. 1978 (rev.). Roadless area-intensive management tradeoffs on western National Forests. Washington, DC: U.S. Department of Agriculture, Forest Service. 57 p.

Hughes, Joy M. 1965. Economic choice and wilderness land allocation. Paper presented to Forestry Section, Northwest Scientific Association Meeting, Portland, OR. 14 p.

Jones, J. Greg; Beardsley, Wendell G.; Countryman, David W.; Schweitzer, Dennis L. 1978. Estimating economic costs of allocating land to wilderness. Forest Science. 24: 410-422.

Kemper, Robert E.; Davis, Lawrence S. 1976. Costs of environmental constraints of timber harvesting and regeneration. Journal of Forestry. 74: 754-761.

Koutsoyiannis, A. 1979. Theory of econometrics. London: The Maximillian Press. 681 p.

MacCleery, Douglas. 1982. Expectations of the Secretary's Office related to National Forest land management planning. Speech given at Regional Foresters and Directors Meeting; 1982 January; Phoenix, AZ. Washington, DC: U.S. Department of Agriculture, Forest Service. 9 p.

Massy, William F. 1965. Principal components regression in exploratory statistical research. Journal of The American Statistical Association. 60: 234-256.

Weisberg, Sanford. 1980. Applied linear regression. New York: John Wiley and Sons. 283 p.

APPENDIX: DEFINITION OF CONSTRAINT VARIABLES USED IN THE PRINCIPAL COMPONENT ANALYSIS CONDUCTED USING FORPLAN SOLUTIONS FROM THE LOLO NATIONAL FOREST

Variable name	Description
PNV	Present net value for a FORPLAN solution measured in thousands of dollars.
CV2	The sum of prescription constraints that specify the thousands of acres to be allocated to deer and elk winter range habitat management without regulated timber, in habitat groups 2, 3, and 6.
CV3	The sum of prescription constraints that specify the thousands of acres to be allocated to grizzly bear habitat management in habitat groups 2, 3, 4, and 5.
CV4	The sum of prescription constraints that specify the thousands of acres to be allocated to old-growth management in habitat groups 2, 3, and 6.
CV5	The sum of prescription constraints that specify the thousands of acres to be allocated to minimum maintenance management on habitat groups 2, 3, 4, and 5.
CV6	The sum of prescription constraints that specify the thousands of acres to be allocated to roadless-dispersed recreation management, including no development or off-road vehicles on habitat groups 0, 2, 3, and 6.
CV7	The sum of prescription constraints that specify the thousands of acres to be allocated to roadless-dispersed recreation management, including no development or off-road vehicles on habitat groups 4 and 5.
CV8	The sum of prescription constraints that specify the thousands of acres to be allocated to a partial retention visual quality objective, with regulated timber on habitat groups 2, 3, and 4.
INVCV1	One divided by the sum of prescription constraints that place an upper limit on the thousands of acres that could be allocated to grizzly bear habitat management on habitat groups 1 and 6.
DWHGEQ	The forestwide sum of prescription constraints that specify the thousands of acres to be allocated to old-growth management.
INVAV4	One divided by the sum of prescription constraints that specify the thousands of acres to be allocated to minimum maintenance management on habitat group 1.
DWVAEQ	The forestwide sum of prescription constraints that specify the thousands of acres to be allocated to regulated timber management, with a retention visual quality objective.
INVKV16	One divided by the sum of prescription constraints that specify the thousands of acres to be allocated to regulated timber and wildlife habitat management having a retention visual quality objective on habitat groups 2 and 3.
MATOLDLL	A lower-limit constraint on the thousands of acre-equivalents of old-growth habitat created for wildlife species that use old-growth.
INVCT21	One divided by the upper limit placed on the amount of timber (units?) that can be clearcut in period 1 on habitat groups 2 and 3.
INVLN19	One divided by the upper limit placed on old-growth management in habitat group 1.
WFORILL	A lower-limit constraint placed on the amount of winter forage for big game (thousands of animal unit months) that is to be produced in decade 1.
HCLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to key elk summer habitat management.
INVNC11	One divided by the sum of prescription constraints that specify the upper limits on the thousands of acres that can be allocated to grizzly bear habitat management in habitat group 4.
HGLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to old-growth management.
RALL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to management for roadless-dispersed recreation in which no development is present and no off-road vehicles are allowed.
NRARB	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to (1) management for roadless-dispersed recreation in which no development is present and no

(con.)

Variable name	Description
	off-road vehicles are allowed, and (2) management for roadless-dispersed recreation in which no timber harvesting is allowed, but off-road vehicles are allowed.
LINEFV18	The sum of all prescription constraints placing a lower limit on the thousands of acres that are to be allocated to management for roadless-dispersed recreation in which no timber harvesting is present and off-road vehicles are allowed in habitat group 5.
RBLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to management for roadless-dispersed recreation in which no development is present and off-road vehicles area allowed.
TBLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to moderate intensity timber management, without local roads, and with soil constraints.
INVNHDL	One divided by the forestwide sum of all upper-limit prescription constraints placed on the thousands of acres to be allocated to grizzly bear habitat management.
INVHCL	One divided by the forestwide sum of all upper-limit prescription constraints placed on the thousands of acres to be allocated to key elk summer range habitat management.
INVNH7	One divided by the sum of all upper-limit prescription constraints placed on the thousands of acres to be allocated to grizzly bear habitat management in habitat group 6.
INVIV19	One divided by the sum of the upper-limit prescription constraints placed on the thousands of acres to be allocated to grizzly bear habitat management in habitat groups 2 and 3.
INVEV19	One divided by the sum of the upper-limit prescription constraints placed on the thousands of acres to be allocated to key elk summer habitat management in habitat group 6.
LINECV8	The sum of prescription constraints placed on the thousands of acres that are to be allocated to key elk summer habitat management in habitat group 4.
LINEEV20	The sum of prescription constraints placed on the thousands of acres that are to be allocated to key elk summer habitat management in habitat group 6.
LINEFV8	The sum of prescription constraints placed on the thousands of acres that are to be allocated to old-growth management in habitat group 5.
LINEAV8	The sum of prescription constraints placed on the thousands of acres that are to be allocated to management for roadless-dispersed recreation in which no development is present and no off-road vehicles are allowed in habitat group 1.
DWRAEQ	The forestwide sum of all prescription constraints placed on the thousands of acres that are to be allocated to management for roadless-dispersed recreation in which no development and no off-road vehicles are allowed.
NHGMM	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to (1) old-growth management, and (2) minimum maintenance management.
NHAVD	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to: (1) Deer and elk winter range habitat management, with regulated timber (2) Deer and elk winter range habitat management, without regulated timber (3) Key elk summer habitat management (4) Grizzly bear habitat management (5) Old-growth management (6) Minimum maintenance management (7) Management for roadless-dispersed recreation in which no development or off-road vehicles are allowed (8) Management for roadless-dispersed recreation in which timber harvesting is not permitted, but off-road vehicles are allowed (9) Management for dispersed recreation in roaded areas (10) Timber management with a retention visual quality objective (11) Timber management with a partial retention visual quality objective (12) Timber and wildlife habitat management with a retention visual management objective (13) Timber and wildlife habitat management with a partial retention visual management objective.

(con.)

APPENDIX (Con.)

Variable name	Description
NHDUL	The forestwide sum of all upper-limit prescription constraints placed on the thousands of acres to be allocated to grizzly bear habitat management.
MMLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to minimum maintenance management.
VCLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to regulated timber and wildlife habitat management having a visual quality objective of retention.
VDLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to regulated timber and wildlife habitat management having a visual quality objective of partial retention.
TRELL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to management for dispersed recreation in roaded areas.
VBLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to regulated timber management having a visual quality objective of partial retention.
NVAVD	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to: (1) Timber management with a retention visual quality objective (2) Timber management with a partial retention visual quality objective (3) Timber and wildlife habitat management with a retention visual management objective (4) Timber and wildlife habitat management with a partial retention visual management objective.
HBLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to deer and elk winter range habitat management without regulated timber.
VALL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to regulated timber management having a visual quality objective of retention.
NHAHB	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to deer and elk winter range habitat management both with and without regulated timber.
LEGHTV4	The forestwide sum of prescription constraints placed on the thousands of acres to be allocated to deer and elk winter range habitat management, with regulated timber in habitat group 1.
NUTBLL	The forestwide sum of all lower-limit prescription constraints placed on the thousands of acres to be allocated to moderate intensity timber management, without local road, and with soil constraints.
NDY2	Dummy variable identifying the presence of the nondeclining yield constraints on timber harvest, coded as: 0: nondeclining yield constraints absent 1: nondeclining yield constraints present.

Hyde, James F. C., III; Jones, J. Greg; Meacham, Mary L. 1988. Evaluating statistical techniques for predicting and interpreting FORPLAN results. Res. Pap. INT-395. Ogden, UT: U.S. Department of Agriculture, Forest Service, Intermountain Research Station. 14 p.

Two approaches using multiple linear regression for analyzing the effects of management constraints on an objective function in FORPLAN were tested on three National Forests. The two approaches, ordinary least squares regression and ordinary least squares using principal components, provide some degree of success in predicting objective function values, but very little information for interpreting the effects of the constraints on objective function values.

KEYWORDS: forest planning, applied regression analysis, land management planning

INTERMOUNTAIN RESEARCH STATION

The Intermountain Research Station provides scientific knowledge and technology to improve management, protection, and use of the forests and rangelands of the Intermountain West. Research is designed to meet the needs of National Forest managers, Federal and State agencies, industry, academic institutions, public and private organizations, and individuals. Results of research are made available through publications, symposia, workshops, training sessions, and personal contacts.

The Intermountain Research Station territory includes Montana, Idaho, Utah, Nevada, and western Wyoming. Eighty-five percent of the lands in the Station area, about 231 million acres, are classified as forest or rangeland. They include grasslands, deserts, shrublands, alpine areas, and forests. They provide fiber for forest industries, minerals and fossil fuels for energy and industrial development, water for domestic and industrial consumption, forage for livestock and wildlife, and recreation opportunities for millions of visitors.

Several Station units conduct research in additional western States, or have missions that are national or international in scope.

Station laboratories are located in:

Boise, Idaho

Bozeman, Montana (in cooperation with Montana State University)

Logan, Utah (in cooperation with Utah State University)

Missoula, Montana (in cooperation with the University of Montana)

Moscow, Idaho (in cooperation with the University of Idaho)

Ogden, Utah

Provo, Utah (in cooperation with Brigham Young University)

Reno, Nevada (in cooperation with the University of Nevada)